ISSN 1684-8403
Journal of Statistics
Volume 14, 2007, pp. 1-12

Most Significant Risk Factors for Head and Neck Cancer

Mueen-ud-Din Azad[1], Ghazia Pervaiz[2] and Muhammad Khalid Pervaiz[3]

Abstract

The Head and Neck Cancer is estimated to be one of the most prevalent cancers in the world. In Pakistan, Head and Neck Cancer accounts for 12.82% of all new cancers registered (1994-2004). The purpose of this research was to investigate the most significant risk factors of Head and Neck Cancer. The risk factors were analyzed in a case-control study of 58 Head and Neck Cancer patients and 130 controls. Patients were recruited over an eight months time period from Institute of Nuclear Medicine and Oncology, Lahore, Pakistan. The Chi-Square test has been used to assess the statistical significant risk factors for the disease of Head and Neck Cancer. The multiple logistic regression model was used to get the most significant risk factors. The study shows that incidence of Head and Neck Cancer is due to five major risk factors: age odds ratio = 1.085 with 95% confidence interval (1.032, 1.140), smoking odds ratio= 99.624 with 95% confidence interval (6.558, 1513.309), chicken odds ratio = 15.997 with 95% confidence interval (1.526, 167.687), beef odds ratio= 13.535 with 95% confidence interval (1.400, 13.840) and soft drinks odds ratio=345.564 with 95% confidence interval (11.870, 10059.919).

Keywords

Odds ratio, Logistic regression, Head and Neck Cancer (HNC)

[1]Department of Statistics, GC University, Lahore
Email: mueen.azad@gmail.com
[2] House Officer, Mayo Hospital, Lahore
Email: ghaziapervaiz @hotmail.com
[3]Department of Statistics, GC University, Lahore
Email: dr.khalidpervaiz@hotmail.com

1. Introduction

Cancer is a group of many related diseases. All cancers begin in living cells which make up blood and other tissues. As these living cells' age, they die and are replaced by new cells. Sometimes, this normal process goes astray; new cells keep forming when the body does not require them and old cells do not die when they should. The mass of extra cells forms a growth or tumor. There are two types of tumors - benign or malignant.

Benign tumors are not cancer, can often be removed and mostly do not come back. Cells from benign tumors do not spread to other parts of the body and are rarely life threatening.

Malignant tumors are cancer. Cells from malignant tumors divide without control and can invade and destroy the tissue around them. Cells also break away from a malignant tumor and enter the bloodstream or lymphatic system. This process, called metastasis, is how cancer spreads from the original (primary) tumor to form new (secondary) tumors in other parts of the body.

Head and Neck Cancers are a diverse group of diseases, each with its own distinct epidemiologic and pathologic features, natural history, and treatment considerations. Despite improvements in diagnosis long-term survival rates for patients with this disease have not increased significantly over the past 30 years and are among the lowest for the major cancers (Monge, 1997). One of the main reasons behind this is that the symptoms are so mild that even with advances in diagnostic modalities physicians do not pick the problem until it is too late, and therapeutic options are, at best, limited (see Table 1).

In the United States, HNC account for 3.2% (39,750) of all new cancers and 2.2% (12,460) of cancer deaths (Wingo et al., 1995). The disease is more common in many developing countries, with a worldwide annual incidence of more than 500,000. As a third world country, Pakistan does not have enough resources to offer treatment to all patients with advanced HNC. By carrying out this study, we aim to identify those risk factors which could play a role in the cause of Head and Neck Cancer.

The incidence of HNC increases with age; most patients are older than age 50. Mainly because of the increasing use of tobacco among women, the male-to-female ratio has decreased from 4 to 5:1 to approximately 3:1 over the past 5 to 10 years (Monge, 1997). Furthermore, some studies have reported a higher risk for women at each successive pack-year stratum of smoking (Spitz et al., 1988), a finding not shown for alcohol (Fracerchi et al., 1994). From 1973 to 1989, the incidence of oral and pharyngeal carcinomas decreased in white men of all ages,

whereas the African-American population experienced a significant increase (Spitz, 1994).

The mucosal surfaces of the upper aerodigestive tract, lungs, and esophagus are repeatedly exposed to the same carcinogens (e.g., tobacco, alcohol). Therefore, multiple independent neoplastic lesions may arise and progress in the same patient, either simultaneously or sequentially, in a process known as "field cancerization" (Slaughter et al., 1953). Metachronous second primary tumors developed at a constant rate of 4% to 7%, are generally of squamous pathology, are not treatment related, and occur in the carcinogen-exposed regions of the aerodigestive tract (Cooper et al., 1989 and Wolf et al., 1993).

A risk factor is anything that increases a person's chance of developing a disease, including cancer. There are risk factors that can be controlled, such as smoking, and risk factors that cannot be controlled, such as age and family history. Although risk factors can influence disease, for many risk factors it is not known whether they actually cause the disease directly. Some people with several risk factors never develop the disease, while others with no known risk factors do. Two risk factors which greatly increase the chance of Head and Neck Cancer (Blum, 2005) are:

- Use of tobacco, including cigarettes, cigars, pipes, chewing tobacco, and snuff (the biggest risk factor for HNC)
- Frequent and heavy consumption of alcohol

Eighty-five percent (85%) of HNC is linked to tobacco use (Blum, 2005). Recent research suggests that people who have used marijuana may be at higher than average risk for HNC (American Society of Clinical Oncology, 2005).

2. Method

In this study, the cases were taken from Institute of Nuclear Medicine and Oncology (INMOL Hospital), Lahore, Pakistan. The cases were taken on the time basis and recruited over eight month's period between January 2005 to August 2005. The controls were selected by using the judgment sampling technique and considering them really healthy. 58 cases and 130 controls were analyzed for the study. A questionnaire technique was used to collect the information from the cases and controls.

The term risk factor is used to designate a variable that is thought to be related to some outcome variable. In the literature we found that gender, age, area

of living, social status, social habits and eating habits are used as the risk factors (Llewellyan et al., 2004; Lee et al., 2004; Redding et al., 1999, and Karabianis et al., 1988). Some new risk factors are also introduced by keeping in view the regional eating and social habits. The risk factors for this research are considered as under:

Gender (GEND), Age (AGE), Marital Status (MARS), Family Size (FAMS), Area of Living (AROL), Education Level (EDUL), Work Place Type (WPT), Monthly Income (MINC), Watching Television (hours per day) (WTV), Car Driving / Traveling (kilometer per day) (CARD), User of Cell Phone (CELP), Smoking (SMOK), Chew Pan (CPAN), Chicken (CHK), Beef (BEEF), Mutton (MUTN), Vegetables (VEG), Pulses (PULS), Oil (OIL), Ghee (GHE), Egg (EGG), Fast Food (FASF), Soft Drink (SOFD), Tea (TEA), Cancer Patient in Family (CPF), Chronic Heart Disease Patient (CHDP), Diabetes Patient (DIBP) and Blood Transfer (BLDT). In further discussion codes will be used. All risk factors are considered as dependent variables.

These risk factors were examined for a univariate association with the HNC. Bivariate analysis of categorical variables was done by the chi-square test or Fisher's exact test. The phi-value has been used to determine the type of association i.e. positive or negative association. Continuous variables were compared by using Student's t-test. Odds Ratio (OR) and 95% confidence intervals (CI) were also calculated. The HNC is taken as dichotomous dependent variable. The probabilities of HNC are calculated by the logistic regression method. The Backward Wald method has been used to obtain the significant risk factors. The final logistic regression model was measured for goodness of fit. The criterion used was Hosmer-Lemeshow test for calibration. A Hosmer-Lemeshow test p-value greater than 0.05 means the fit is adequate. The Wald's statistic has been used to test the significance of individual logistic regression coefficients for each risk factor. The null hypothesis that a particular logit coefficient is zero has been tested. The factors with insignificant effect were dropped from the multiple logistic model. All statistical analyses have been performed by using SPSS, version 13.0 (SPSS, Chicago- III).

3. Results

A comparison of the variables used for predicting the presence of HNC between cases and controls is done by using the descriptive statistics presented in Table 2. The cases and controls were between 04 and 80 year of age with a mean 39.69 years (CI = 39.69 $\pm$ 16.01). The family size lies between 0 and 12 persons with a

mean 4.74 persons (CI= 4.74 ± 2.45). The duration of watching television lies between 0 to 8 hours per day with mean 2.07 hours (CI=2.07 ± 1.71). The traveling lies between 0 to 150 km per day with mean 13.33 km per day (CI=13.33 ± 22.43).

The variables that are statistically significant and have positive association with the presence of HNC were: AGE, MARS, SMOK, CHK, BEEF, GHE, SOFD and BLDT. It shows that as the above variables quantity increases, the risk of HNC increases (see Table 3).

The variables that are statistically significant and have negative association with the presence of HNC on univariate analysis were: GEND, AROL, WTV, MINC, CELP, MUTN, VEG, PULS, EGG, FASF, OIL, and TEA. It shows that as the above mentioned variables increase the risk of HNC decreases (see Table 3).

Using multiple logistic regression analysis all the 29 variables entered into the linear predictor. The final set of variables was selected by the Backward Wald method. The fit of this model to the data was excellent. The Hosmer-Lemeshow statistic for goodness of fit was 0.232 with p-value 0.99 showing that the model adequately fits the data. One hundred and twenty five out of 130 (96.2 %) persons without HNC are correctly predicted by the model. Similarly, 53 out of 58 (91.4 %) patients with HNC were correctly predicted.

The p-values of ten factors - age, monthly income, cell phone, smoking, chicken, beef, mutton, oil, egg and soft drink - have significant contribution. All the remaining factors appear as insignificant. The odds ratios for the variables along with coefficients are given in Table 4.

The first factor is age; it attains numeric value. The coefficient of age has positive value 0.081. The odds ratio is 1.09 with 95% confidence interval (1.03, 1.14). It means that with the increase of one year in the age the risk of HNC is increased 1.085 times provided all other factors are kept constant.

The second factor is income. We coded the income into two categories: one that has monthly income less than Rs. 7,000/- with coded value 0 and the above or equal to with coded value 1. The coefficient of income has negative value -4.193 with odds ratio 0.015 with 95% confidence interval (0.001, 0.247). This shows that as the income status goes up, the chance of HNC goes down.

Third factor is cell phone. The coefficient of cell phone is negative which is -6.697. The odds ratio is 0.001 with 95% confidence interval (0.00, 0.12). This shows that it is a negatively significant risk factor with HNC.

The fourth significant factor is smoking. The coefficient of smoking is 4.601 which is positive. The odds ratio is 99.62 with 95% confidence interval

(6.56, 1513.31). This shows high significance with HNC. It also shows that a person who is smoker has 99 times more chances to be a HNC.

The fifth factor is chicken. The coefficient of chicken is 2.772 and is positive. The odds ratio is 16.00 with 95% confidence interval (1.53, 167.69). This shows high significance with HNC. This also shows that a person who eats chicken regularly has 16 times more chances to have HNC.

The sixth factor is beef. The coefficient of beef is 2.605 and is positive. The odds ratio is 13.54 with 95% confidence Interval (1.40, 130.84). This shows high significance with HNC. This shows that a person who eats beef has 13 times more chances to be a HNC.

The seventh factor is mutton. The coefficient of mutton is -4.153 and is negative. The odds ratio is 0.016 with 95% confidence interval (0.001, 0.20). This shows that a person who eats mutton has about 98.4 % chance to be healthy.

The eighth factor is eggs. The coefficient of eggs is -4.143 and is negative. The odds ratio is 0.016 with 95% confidence interval (0.001, 0.20). This also shows that a person who eats eggs has about 98.4 % chance to be healthy.

The ninth factor is cooking oil. The coefficient of oil is -2.784 and is negative. The odds ratio is 0.062 with 95% confidence interval (0.01, 0.37). This also shows that a person who uses cooking oil has about 93.8 % chance to be healthy.

The tenth factor is soft drink. The coefficient of soft drink is 5.845 and is positive. The odds ratio is 345.564 with 95% confidence interval (11.87, 10059.92) greater than 1. This shows high significance with HNC. It also shows that a person who is fond of soft drink has a 346 times more chances to be a HNC. The final fitted logistic model is as under:

$$\hat{y} = p(HNC) = \frac{1}{1+e^{-z}}$$

where

$$Z = -1.872 + 0.081(AGE) - 4.193(MINC) - 6.697(CELP) + 4.601(SMOK) + 2.772(CHK) + 2.605(BEEF) - 4.153(MUTN) - 4.143(EGG) - 2.784(OIL) + 5.845(SOFD)$$

4. Discussion

The main aim of this case control study was to evaluate and dig out the key risk factors for HNC. The age is an uncontrollable variable. In these cases, 20 (34.48%) were above the age 50 years. The odds ratio for age is 1.09. It means

that with the increase of one year in the age, the risk of HNC is increased 1.085 times provided all other factors are kept constant.

In these cases, 35 (60.34%) were male and the rest 23 (39.66%) were females. This shows that the number of males is higher for HNC, as reported by Eaden et al. (2000).

Only the four cases used the cell phone. The negative coefficient (-0.70) shows that the effect of cell phone is positive, that is, the use of cell phone decreases the chance of HNC. In literature no such relationship is found. This may be due to very small number of cases using the cell phone.

In these cases, 37 (63.79%) were nonsmokers and the rest 21 (36.21%) were smokers. The odds ratio for smoking is 99.63 which shows that the chance of incidence of HNC in smokers is 99 times more than in nonsmokers. This finding also matches the previous work as mentioned by Eaden et al. (2000), Lee et al. (2004) and Llewellyn et al. (2004).

The twenty-five (43.10%) cases eat chicken regularly. The phi-value for chicken is 0.15, which shows positive association. The coefficient of chicken is also positive and the odds ratio is 16.0; therefore, by using the chicken regularly the risk of HNC increases by 16 times.

The thirty-two (55.17%) cases eat beef regularly. The p-value is 0.48, which indicates the positive association. The coefficient of beef is also positive and the odds ratio is 13.54; therefore, by using the beef regularly the risk of HNC increases by 14 times.

In the cases, 40 (68.97%) use the soft drinks regularly. The phi-value is 0.16, which indicates the positive association. The coefficient of soft drink is also positive and the odds ratio for the soft drink is 345.57. Hence by using the soft drink the risk of HNC increases by 346 times.

The results suggested that the occurrence of HNC is related to age, smoking and use of chicken, beef and soft drinks regularly.

5. Limitations of the study

In this study, we did not look for the use of Alcohol and Betel Quid among our subjects, even though these two are significant risk factors in the development of HNC. The following is our justification for each:

1- ALCOHOL: Pakistan is a Muslim majority country. Muslim Law prohibits the trade as well as consumption of alcohol. Therefore, an insignificant section of our population uses alcohol.
2- BETEL QUID: Though common in South-East Asia, the use of Betel Quid

is very rare in urban Punjab, the target population of this study. Therefore, we did not consider this factor significant enough to be included in our study.

Table1: Basic Information about Site and Symptoms of Head and Neck Cancer

Type of Cancer (According to body area involved)	**Description**	**Symptoms**
Oral Cavity	Lips, front 2/3 of tongue, gums, lining inside cheeks & lips, part of mouth beneath the tongue, the top, bony part of mouth (hard palate) & the small area behind wisdom teeth	White or red patch on gums, tongue, or lining of the mouth, swelling of jaw, unusual bleeding, pain in the mouth
Nasal Cavity & Sinuses	Sinuses are hollow cavities in bones around the nose	Chronically blocked and infected sinuses that are unresponsive to treatment
	Nasal cavity is a hollow space on each side of the nose	Bleeding through the nose, frequent headaches, pain in upper teeth, problems with dentures, swelling of eyes
Salivary Glands	Present in mouth. Produce saliva, which keeps the mouth moist and aids in chewing food	Swelling under chin or around jawbone, numbness or paralysis of face muscles, unremitting pain in face, chin & neck
Pharynx	Hollow tube that runs between nose and esophagus and the trachea (the tubes connected to the stomach and lungs, respectively)	Trouble in breathing or speaking, headaches, pain or ringing in ears, hearing problems
Larynx	Responsible for producing voice & preventing food from entering air passages during swallowing	Pain on swallowing, hoarseness of voice, ear pain
Lymph Nodes	Traditionally site of cancer spread, but, occasionally, squamous cell cancer found without any evidence of primary cancer in the Head & Neck area.	Pain or swelling in the neck or throat that does not go away

(National Cancer Institute, US National Institutes of Health)

Table 2: Characteristics of Cases and Controls

	Controls	**Cases**	**Total**		**Controls**	**Cases**	**Total**
GENDER				EAT VEGETABLE REGULARLY			
Female	46	23	69	No	13	25	38
Male	84	35	119	Yes	117	33	150
Total	130	58	188	Total	130	58	188
MARITAL STATUS				EAT PULSES REGULARLY			
Single	38	8	46	No	60	48	108
Married	92	50	142	Yes	70	10	80
Total	130	58	188	Total	130	58	188
AREA OF LIVING				EAT EGG REGULARLY			
Rural	53	41	94	No	18	23	41
Urban	77	17	94	Yes	112	35	147
Total	130	58	188	Total	130	58	188
MONTHLY INCOME				EAT FAST FOOD			
> 7000	59	55	59	NO	110	55	165
≤7000	71	3	71	YES	20	3	23
Total	130	58	130	Total	130	58	188
CELL PHONE				USE COOKING OIL			
No	43	54	97	No	28	44	72
Yes	87	4	91	Yes	102	14	116
Total	130	58	188	Total	130	58	188
SMOKEING				USE GHEE			
No	112	37	149	No	102	14	116
Yes	18	21	39	Yes	28	44	72
Total	130	58	188	Total	130	58	188
EAT CHICKEN REGULARLY				TAKE SOFT DRINK			
No	53	33	86	No	63	18	81
Yes	77	25	102	Yes	67	40	107
Total	130	58	188	Total	130	58	188
EAT BEEF				TAKE TEA			
No	116	26	142	No	10	30	40
Yes	14	32	46	Yes	120	28	148
Total	130	58	188	Total	130	58	188
EAT MUTTON REGULARLY							
No	32	47	79				
Yes	98	11	109				
Total	130	58	188				

Table 3: Results of Chi-Square Test

Risk Factors	**Pearson Chi-Square**	**d. f.**	**p-value**	**phi-value**
Marital Status (MARS)	5.172	1	0.023	0.166
Smoking (SMOK)	12.197	1	0.000	0.255
Chicken (CHK)	4.203	1	0.040	0.150
Beef (BEEF)	42.787	1	0.000	0.477
Ghee (GHE)	50.086	1	0.000	0.516
Soft Drink (SOFD)	4.967	1	0.026	0.163
Blood Transfer (BLDT)	11.071	1	0.001	0.243
Area of Living (AROL)	14.362	1	0.000	-0.276
Work Place Type (WPT)	22.967	1	0.000	-0.350
Monthly Income (MINC)	41.077	1	0.000	-0.467
Cell Phone (CELP)	57.863	1	0.000	-0.555
Mutton (MUTN)	52.400	1	0.000	-0.528
Vegetables (VEG)	27.252	1	0.000	-0.381
Pulses (PULS)	21.983	1	0.000	-0.342
Egg (EGG)	15.666	1	0.000	-0.289
Fast Food (FASF)	3.895	1	0.048	-0.144
Cooking Oil (OIL)	50.086	1	0.000	-0.516
Tea (TEA)	46.424	1	0.000	-0.497

Table 4: Significant Predictors of HNC

	B	Wald	d f.	p-value	Odds Ratio	95% C.I.	
						Lower	Upper
AGE	0.081	10.223	1	0.001	1.085	1.032	1.140
MINC(1)	-4.193	8.646	1	0.003	0.015	0.001	0.247
CELP(1)	-6.697	8.292	1	0.004	0.001	0.000	0.118
SMOK(1)	4.601	10.988	1	0.001	99.624	6.558	1513.309
CHK(1)	2.772	5.348	1	0.021	15.997	1.526	167.687
BEEF(1)	2.605	5.066	1	0.024	13.535	1.400	130.840
MUTN(1)	-4.153	10.078	1	0.002	0.016	0.001	0.204
EGG(1)	-4.143	10.306	1	0.001	0.016	0.001	0.199
OIL(1)	-2.784	9.334	1	0.002	0.062	0.010	0.369
SOFD(1)	5.845	11.549	1	0.001	345.564	11.870	10059.919
CONSTANT	-1.872	1.738	1	0.187	0.154		

Acknowledgements

We would like to thank Mr. Mumtaz-ul-Haq (Head, Medical Physics Department, INMOL) and Mr. Khalid Masood (Senior Scientist, INMOL), who kindly gave permission for access to cancer patient records.

References

1. American Society of Clinical Oncology (2005). http://www.plwc.org/portal/site/PLWC/menuitem.
2. Blum (2005). http://www.plwc.org
3. Cooper, J. S, Pajak, T. F., Rubin, P. Tupchong, L., Brady, L. W., Leibel, S. A., Laramore, G. E., Marcial, V. A., Davis, L.W. and Cox. J. D. (1989). Second malignancies in patients who have head and neck cancer: incidence, effect on survival and implications based on the RTOG experience. *International Journal of Radiation Oncology Biology Physics*, **17(3)**, 449–456.
4. Eaden, J., Abrams, K., Ekbom, A., Jackson, E. and Mayberry, J. (2000). Colorectal cancer prevention in ulcerative colitis: a case-control study. *Alimentary Pharmacology and Therapeutics*, **14(2)**, 145-153.
5. Fracerchi, S., Bidoli, E., Negri, E., Barbone, F. and Vecchia, C. L. (1994). Alcohol and cancer of the upper aerodigestive tract in men and women. *Cancer Epidemiology Biomarkers & Prevention*, **3(4)**, 299–304.
6. Karabianis, A., Hill, C., Leclereq, B., Tancrede, C., Baume, D. and Andremont, A. (1988). Risk factors for candidemia in cancer patients: a case-control study. *Journal of Clinical Microbiology*, **26 (3)**, 429-432.
7. Lee, S.-O., Kim, N. J., Choi, S.-H., Kim, T. H., Chung, J.W., Woo, J. H., Ryu, J. and Kim, Y. S. (2004). Risk factors for acquisition of imipenem-resistant *Acinetobacter Baumannii*: a case-control study. *Antimicrobial Agents and Chemotherapy*, **48(1)**, 224–228.
8. Llewellyn, C. D., Jhonson, L. W. and Warnakulasuriya, K. A. (2004). Risk factors for oral cancer in newly diagnosed patients aged 45 years and younger: a case-control study in Southern England. *Journal of Oral Pathology Medicine*, **33(9)**, 525-532.
9. Monge, E. J. (1997). "MEDICAL ONCOLOGY: A COMPREHENSIVE REVIEW" http://www.cancernetwork.com/textbook/morev13.htm
10. National Cancer Institute, US National Institutes of Health http://www.cancer.gov/cancertopics/factsheet/sites-types/head-and-neck

11. Redding, S. W., Zellars, R. C., Kirkpatrick, W. R., McAtee, R. K., Caceres, M. A., Fothergill, A. W., Lopez-Ribot, J. L., Bailey, C. W., Rinaldi, M. G. and Patterson, T. F. (1999). Epidemiology of oropharyngeal candiada colonization and infection in patients receiving radiation for head and neck cancer. *Journal of Clinical Microbiology*, **37(12)**, 3896-3900.
12. Slaughter, D. L., Southwick, H. W. and Smejkal, W. (1953). Field Cancerization in Oral Stratified Squamous Epithelium: Clinical Implication of Multicentric Origin. *Cancer*, **6**, 963–968.
13. Spitz, M. R. (1994). Epidemiology and Risk Factors for Head and Neck Cancer. *Oncology*, **21**, 281–288.
14. Spitz, M. R., Fueger, J. J., Goepfert, H., Hong, W. K. and Newell, G.R. (1988). Squamous cell carcinoma of the upper aerodigestive tract: a case comparison analysis. *Cancer*, 61, 203–208.
15. Wingo, P., Tony, T. and Bolden, S. (1995). Cancer Statistics. *CA: A Cancer Journal for Clinician*, **45**, 8–30.
16. Wolf, G., Lippman, S. M. and Laramore, G. (1993). Head and Neck Cancer in Holland. *Cancer Medicine*, **3**, 1211–1278.